

This listing of claims will replace all prior versions, and listings, of claims in the application.

**Listing of Claims:**

1. (Previously presented) A method for selecting a server from a plurality of servers to service a third request for an asset, comprising:
  - generating, at an adaptable cache of a first server, a list of pairs of requests for the asset and determining a pair of requests having the shortest interval between start times, wherein the pair of requests comprise a first request and a second request;
  - responsive to the first request, streaming the asset from a storage system of the first server and storing the asset on the adaptable cache of the first server as it is being streamed from the storage system of the first server;
  - responsive to the second request, streaming the asset from the adaptable cache of the first server;
  - updating a first state table on the first server with information about the asset stored on the adaptable cache of the first server;
  - communicating the information about the asset stored on the adaptable cache of the first server to each server in the plurality of servers;
  - updating state tables of each of the other servers in the plurality of servers with the information about the asset stored on the adaptable cache of the first server;
  - designating a director from the plurality of servers to receive the third request, wherein any of the plurality of servers can be designated as the director;
  - determining that the asset is not stored on an adaptable cache of the director by accessing a director's state table stored on the director, wherein the director's state table includes parametric information for each server in the plurality of servers, and wherein the parametric information comprises adaptable cache contents information for each server in the plurality of servers; and
  - under the direction of the director,
    - determining a set of servers from among said plurality of servers that have the asset stored in their respective adaptable caches by examining the state table on the director;
    - determining a load factor for each of the set of servers; and

selecting a second server from among the set of servers based on the load factor.

2. (Currently amended) The method of claim 1, wherein ~~the step of designating the~~ director comprises designating the director in a round-robin fashion.

3. (Previously presented) The method of claim 1, wherein the director is designated based on a load factor analysis for each server among said plurality of servers, the load factor for each server based on parametric information stored in a respective state table thereon, and wherein the designated director has a lowest load factor.

4. (Currently amended) The method of claim 1, further comprising ~~selecting~~ designating the director upon determining that the asset is present on the director.

5. (Previously presented) The method of claim 1, wherein said parametric information further comprises functional state and current load of each server.

6. (Previously presented) The method of claim 1, wherein said parametric information further comprises whether each server comprises extended memory.

7. (Previously presented) The method of claim 1, wherein said parametric information further comprises whether each server comprises an inline adaptable cache.

8. (Previously presented) The method of claim 1, wherein said parametric information further comprises whether each asset represented in the parametric information is a new release.

9. (Previously presented) The method of claim 1, further comprising storing the asset on the adaptable cache of the director responsive to the third request.

10. (Previously presented) The method of claim 1, further comprising forwarding the third request to the second server.

11. (Previously presented) The method of claim 1, further comprising redirecting the third request to the second server.

12. (Previously presented) The method of claim 1, wherein selecting the second server from among the set of servers further comprises:

identifying as available servers any servers whose load factors are below threshold limits;

determining that there are no available servers; and

upon determining that there are no available servers, selecting a server having a lowest load factor from the other servers having the content.

13-20 (Cancelled)

21. (Currently amended) A computer-readable medium comprising computer-executable instructions for performing a method comprising:

generating, at an adaptable cache of a first server, a list of pairs of requests for an asset and determining a pair of requests having the shortest interval between start times, wherein the pair of requests comprise a first request and a second request;

responsive to the first request, streaming the asset from a storage system of the first server and storing the asset on the adaptable cache of the first server as it is being streamed from the storage system of the first server;

responsive to the second request, streaming the asset from the adaptable cache of the first server;

updating a first state table on the first server with information about the ~~the~~ asset stored on the adaptable cache of the first server, wherein the first server communicates the information about the asset stored on the adaptable cache of the first server to each server in the plurality of servers, and wherein each server in the

plurality of servers updates each state table of each server in the plurality of servers with the information about the asset stored on the adaptable cache of the first server;

designating a director from the plurality of servers to receive a third request, wherein any of the plurality of servers can be designated as the director;

determining that the asset is not stored on an adaptable cache of the director by accessing a state table stored on the director, wherein the state table includes parametric information for each server in the plurality of servers, and wherein the parametric information comprises adaptable cache contents information for each server in the plurality of servers; and

under the direction of the director,

determining a set of servers from among said plurality of servers that have the asset stored in their respective adaptable caches by examining the state table on the director;

determining a load factor for each of the set of servers; and,

selecting a second server from among the set of servers based on the load factor.

22. (Currently amended) The computer-readable medium of claim 21, wherein ~~the step of designating~~ the director comprises designating the director in a round-robin fashion.

23. (Currently amended) The computer-readable medium of claim 21, wherein ~~the step of designating~~ the director comprises designating the director on the basis of lowest load.

24. (Currently amended) The computer-readable medium of claim 21, wherein ~~the step of selecting~~ designating the director further comprises ~~selecting~~ designating the director if the requested content is present on the director.

25. (Previously presented) The computer-readable medium of claim 21, wherein said parametric information further comprises functional state and current load of each server.

26. (Previously presented) The computer-readable medium of claim 21, wherein said parametric information further comprises whether each server comprises extended memory.

27. (Previously presented) The computer-readable medium of claim 21, wherein said parametric information further comprises whether each server comprises an inline adaptable cache.

28. (Previously presented) The computer-readable medium of claim 21, wherein said parametric information further comprises whether each asset represented in the parametric information is a new release.

29. (Previously presented) The computer-readable medium of claim 21, further comprising computer-executable instructions for storing the asset on the adaptable cache of the director responsive to the third request.

30. (Previously presented) The computer-readable medium of claim 21, further comprising computer-executable instructions for forwarding the third request to the second server.

31. (Previously presented) The computer-readable medium of claim 21, further comprising computer-executable instructions for redirecting the third request to the second server.

32. (Previously presented) The computer-readable medium of claim 21, wherein the step of selecting the second server from among the set of servers further comprises:

- identifying as available servers one or more servers whose load factors are below threshold limits;
- determining that there are no available servers; and
- upon determining that there are no available servers, selecting a server having a lowest load factor from the other servers having the content.

33. (Currently amended) The method of claim 1, further comprising updating parametric information in a second state table ~~associated with the selected~~ on the second server, and communicating updated parametric information to the other servers among said plurality of servers.

34. (Previously presented) The method of claim 33, wherein the updated parametric information is communicated via multicast.

35. (Previously presented) The method of claim 33, wherein the updated parametric information is communicated via a broadcast message.

36. (New) A system for servicing a request for an asset, comprising:

- an adaptable cache of a first server configured to generate a list of pairs of requests for the asset and determine a pair of requests having the shortest interval between start times, wherein the pair of requests comprise a first request and a second request;

- a storage system of the first server configured to stream the asset responsive to the first request;

- the adaptable cache of the first server further configured to store the asset as the asset is streamed from the storage system of the first server;

- the adaptable cache of the first server configured to stream the asset responsive to the second request;

- the first server configured to update a first state table with information about the asset stored on the adaptable cache of the first server;

- the first server further configured to communicate the information about the asset stored on the adaptable cache of the first server to each server in the plurality of servers;

- each server in the plurality of servers configured to update their respective state tables with the information about the asset stored on the adaptable cache of the first server;

a business management system configured to designate a director from the plurality of servers to receive the third request, wherein any of the plurality of servers can be designated as the director;

the director configured to determine that the asset is not stored on an adaptable cache of the director by accessing a director's state table, wherein the director's state table includes parametric information for each server in the plurality of servers, and wherein the parametric information comprises adaptable cache contents information for each server in the plurality of servers; and

the director further configured to:

determine a set of servers from among said plurality of servers that have the asset stored in their respective adaptable caches by examining the director's state table,

determine a load factor for each of the set of servers; and

select a second server from among the set of servers based on the load factor.

37. (New) The system of claim 36, wherein the business management system configured to designate the director comprises the business management system configured to designate the director in a round-robin fashion.

38. (New) The system of claim 36, wherein the business management system configured to designate the director comprises the business management system configured to designate the director based on a load factor analysis for each server among the plurality of servers, the load factor for each server based on parametric information stored in a respective state table thereon, and wherein the designated director has a lowest load factor.

39. (New) The system of claim 36, wherein the business management system configured to designate the director comprises the business management system configured to designate the director upon determining that the asset is present on the director.

40. (New) The system of claim 36, wherein said parametric information further comprises at least one of a functional state of each server, a current load of each server, whether each server comprises extended memory, whether each server comprises an inline adaptable cache, and whether each asset represented in the parametric information is a new release.

41. (New) The system of claim 36, wherein the director is further configured to perform at least one of storing the asset on the adaptable cache of the director responsive to the third request, forwarding the third request to the second server, and redirecting the third request to the second server.

42. (New) The system of claim 36, wherein the director configured to select the second server from among the set of servers further comprises the director configured to:

- identify as available servers any servers whose load factors are below threshold limits;
- determine that there are no available servers; and
- upon determining that there are no available servers, select a server having a lowest load factor from the other servers having the content.

43. (New) The system of claim 36, further comprising updating parametric information in a second state table on the second server, and communicating updated parametric information to the other servers among said plurality of servers.

44. (New) A system for servicing a request for an asset, comprising:

- an adaptable cache means of a first server for generating a list of pairs of requests for the asset and determining a pair of requests having the shortest interval between start times, wherein the pair of requests comprise a first request and a second request;
- means for streaming, responsive to the first request, the asset from a storage system of the first server and storing the asset on the adaptable cache means of the first server as it is being streamed from the storage system of the first server;



means for streaming the asset from the adaptable cache means of the first server responsive to the second request;

means for updating a first state table on the first server with information about the asset stored on the adaptable cache means of the first server;

means for communicating the information about the asset stored on the adaptable cache means of the first server to each server in the plurality of servers;

means for updating state tables of each of the other servers in the plurality of servers with the information about the asset stored on the adaptable cache means of the first server;

means for designating a director from the plurality of servers to receive the third request, wherein any of the plurality of servers can be designated as the director;

means for determining that the asset is not stored on an adaptable cache of the director by accessing a director's state table stored on the director, wherein the director's state table includes parametric information for each server in the plurality of servers, and wherein the parametric information comprises adaptable cache contents information for each server in the plurality of servers; and

means under the direction of the director for:

determining a set of servers from among said plurality of servers that have the asset stored in their respective adaptable caches by examining the state table on the director;

determining a load factor for each of the set of servers; and

selecting a second server from among the set of servers based on the load factor.